Procedia Computer Science
Volume 51, 2015, Pages 2117–2126
ICCS 2015 International Conference On Computational Science

The Resilience of the Encounter Network of Commuters for a Metropolitan Public Bus System.

Muhamad Azfar Ramli*and Christopher Pineda Monterola

Institute of High Performance Computing,
Agency for Science, Technology and Research,
Fusionopolis, 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632
ramlimab@ihpc.a-star.edu.sg,monterolac@ihpc.a-star.edu.sg

Abstract

We analyse the structure and resilience of a massive encounter network generated from commuters who share the same bus ride on a single day within a metropolitan city. We demonstrate our analysis using a case study of the network created by all the commuters who utilised the public bus system during a typical weekday in the whole of Singapore using smartcard data. We show that the network structure is of random-exponential type with small world features rather than a scale-free network. Within one day, 99.97% of all commuters are connected approximately within 7 steps of each other. We report on how this network structure changes upon application of a threshold based on the encounter duration (T_E). Among others, we demonstrate that the total number of connected commuters reduces by 50% when we set the threshold for T_E to be 15mins. We then assess the dynamics of infection spreading by comparing the effect of both random and targeted removal strategies of commuters. By assuming that the network characteristic is invariant day after day, our simulation indicates that without node removal, 99% of the commuter network will be infected within 7 days of the onset of infection. While a targeted removal strategy is demonstrated to delay the onset of the maximum number of infected individuals, it is not able to effectively isolate commuters away from being eventually infected.

Keywords: epidemic spread, commuter network, network attack, public transport,

1 Introduction

The global population of urban metropolitan cities has grown rapidly for the past decade and is projected to continue on an upward trajectory throughout the century. The threat of a serious global epidemic in the modern world has also become closer to reality with the promulgation of deadly diseases such as avian influenza, Ebola and SARS. Understanding the mechanism

*Corresponding author

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2015

doi:10.1016/j.procs.2015.05.482

of spread of infectious diseases in these highly clustered city landscapes is therefore a major concern for all city planners around the world.

Previous studies have looked at the epidemic spreading from a structural perspective and found that the network structure is critical to the mechanism of epidemic spread. Most notably, scale-free networks have shown to have distinct spreading characteristics as compared to traditional random networks [7], especially the absence of an epidemic threshold [11]. Various other studies utilize synthetic network generation methods to highlight the different spreading characteristics [5] while studies with actual empirical data are much harder to come by.

Spreading dynamics on online social networks are commonly analysed due to the ease of mining their network datasets [12]. These studies also have interesting applications such as curbing the spreading of computer viruses [11] and information transmission within the online realm. Social networks typically have scale-free properties and significantly long-tailed degree distributions as the total number of connections have no upper band constraints. However, physical contact networks have different structural properties as compared to online systems due to physical limitations to the amount of contact people might have with each other.

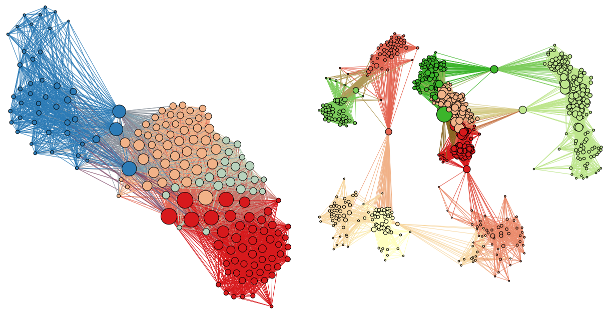
Using empirical data, the effect of the rate of mobility [8] on the spreading of diseases have been studied both for the case of intra-city [6][17] as well as the longer range inter-city [3][2] travel. However, few of these methods are granular enough to account for individual-level interactions. Most notably we note the work of Sun *et al* [15] who previously investigated the encounter network created from commuters in the public bus system of Singapore. An investigation was also conducted on the susceptibility of the network to epidemic spread by using a simulation based approach using a variant of the Susceptible-Exposed-Infected Model (SEI). The investigation also recommended a methodology for early detection of a major outbreak using the highest connected individuals as detectors [14].

In this article, we analyse the structure of the bus encounter network of a typical metropolitan city with the aim of identifying effective strategies to mitigate epidemic spreading. This is conducted by considering the resilience of the network which has previously been found to be highly dependent on the network structure [1]. We follow the infection spreading simulation conducted in [14] but with a recovery state (SIR) in order to make a more realistic assessment on the potential maximum number of infected persons in the network. We then assess the impact of various node removal strategies on the dynamics of infection spread anchored on the empirically measured transmissivity rate of virulent influenza.

2 Building the Massive Commuter Encounter Network of a Metropolitan Public Bus System

We take the public transportation network of Singapore as a case study for understanding the nature and dynamics of encounters that occur between different commuters. In Singapore, the public bus system together with the Mass Rapid Transit (MRT) and Light Rail Transit (LRT) Systems form the primary modes of public transportation. Whilst the MRT is the primary mode of transportation of choice for journeys involving cross-island travel, public bus services serve both as an alternative to MRT for making long range travel and intra-neighborhood travel. The bus system is serviced by approximately 4600 bus stops spread across the island and consists of about 340 different bus services varying in route length and distance [10]. Whilst it is difficult to study commuter encounters within the MRT and LRT systems due to lack of granular data, encounters occurring within the public bus system is easily discerned from smartcard data.

In order to build a network model that represents all the encounters between the various



(a) Commuter network for one journey (b) Multiple connected journeys

Figure 1: A visualization of a sample encounter network of commuters for a) a single bus trip b) several journeys of the same bus linked together. The size of the node is proportional to the degree while nodes of a similar colour and grouped by the time of the journey. Note that there are several commuters which link different clusters together ensuring full connectivity of the network. Network visualizations are created using the open source software Gephi [4].

commuters, we utilised an anonymised historical dataset obtained from the Land Transport Authority of Singapore that records about 3.4 million commuter journeys consisting of about 1.5 million unique commuters who utilised the public bus service for a single day in March 2012 within Singapore. When a commuter boards a bus during its journey by tapping his smartcard onto the reader, the dataset stores the record of the unique ID of the smartcard, the bus ID, the time, as well as the starting bus stop location. The same information is also recorded at the end of the journey when the commuter taps his card to leave the bus. By combining the information of different commuters who were together on the same bus ID during the same journey, we can discern which passengers encountered each other during the course of their journey. Each commuter that boarded a bus is thereby represented by a node in the network model. When he boards the bus, a network edge is added between him and every commuter who was already present in the bus. When the commuter leaves the bus, the different amounts of time spent on the bus together with the remaining commuters is recorded as an edge attribute in the network. We refer to this time as the encounter duration T_E throughout this paper. In cases where multiple edges are formed between two individuals on the same day, we simply take the total encounter duration given by the sum of all encounter times happening within the day.

The networks of all the journeys of the 3617 active buses for the day are then combined together into a single massive nationwide network. Notably, selected passengers would unwittingly link the different networks together whenever they travel on multiple journeys within the day (refer to Fig. 1(b)). We also note that the mean number of journeys per commuter obtained from the dataset is 2.27. This number is indeed consistent with the average number of two journeys per commuter, which typically consists of the morning work/school commute followed by the commute home in the evening.

Network Statistics	Quantity
Total number of nodes (N)	1,490,565
Number of isolated nodes	294
Total number of edges (E)	75,670,986
Network Density ($\frac{2E}{N(N-1)}$)	6.81×10^{-5}
Mean degree $\bar{k}_n$	101.55
Mean Encounter Duration (T_E)	9.64 mins
Clustering Coefficient	0.56
Size of GC (% of N)	1,490,140 (99.97%)
Diameter of GC (approx.)	7

Table 1: Summary of statistics for the commuter encounter network

3 Features of the massive commuter encounter network

Here we provide several definitions of network statistics that we use as measures to characterise the encounter network. The size of the network denoted as N is the total number of nodes (n) in the network which corresponds to the number of commuters who were recorded to have utilized the bus in the smartcard dataset. The total number of edges denoted as E corresponds to all of the encounters which occurred between any two commuters and the number of edges adjacent to any node is known as the degree of the node n and this is denoted as k_n. The encounter duration T_E is defined earlier as the total amount of time spent between two commuters during their encounters for the day.

A connected component of the network is any subset of the network in which all of its nodes can be reached in a finite number of steps. The size of the giant connected component (we refer to this as GC throughout this article) is the maximum size out of all the connected components obtained from the network. This provides a measure of the potential reach of any infectious disease within the network. Given an unlimited amount of time, an infected node that is found within the GC will essentially be able to infect all of the nodes within itself.Another network measure is the diameter of the network defined as the maximum shortest path between any two nodes in the network. This measure represents the capability or reach of the network and a smaller diameter can be associated with faster transmission between nodes.

Table 1 summarizes of the basic characteristics and statistics of the network obtained from the encounters of the commuters in the dataset. We observe that this massive network appears to be highly clustered (a random Erdos-Renyi network of comparable size yielded a clustering coefficient of 6.78×10^{-5} as compared to the 0.56 obtained here) with a significantly high mean degree of about 100. We note that this number is indeed consistent with the capacity of the buses used in the Singapore bus system which is 90 for a single-decker and 133 for a double decker. The network is expected to have high clustering since all possible edges are created between every commuter that was present together in the bus.

Notably, the degree distribution of the network is found to be exponentially distributed (refer to Fig. 2a) and not scale-free, this can be explained by understanding that it would not be physically possible for a number of individuals to have a exponentially large number of edges due to physical contraints on the vehicle size. This result raises major implications on our attempts to characterise the resilience of the network since scale-free networks and exponential networks have been shown to respond very differently to different mechanisms of attack [1]. Discussion on this concern is covered in more detail in Section 5.

The resulting GC for our encounter network is found to consist of 99.97% of the total number of nodes in the network. In addition, the approximate diameter of this 1.5 million node network is only 7 showing that the network exhibits the small-world property [16]. While this result is expected for random networks with a high degree of connections, the result is still alarming as it implies that theoretically a single commuter with a highly infectious disease will essentially be able to infect almost all of the 1.5million commuters in the city within 7 steps.

4 Effect of thresholding encounter duration on the network

The capability of the encounter in transmitting an infectious disease is dependent on the amount of time spent during the encounter [9]. Figure 2 shows the impact of applying a threshold based on the encounter duration T_E suggesting that:

$$f(k_n) \approx a \exp^{-bk_n} \text{ where } a, b \approx g(T_E) \tag{1}$$

The exponent b gradually decreases with T_E up to a threshold of 25mins until which it remains stabilizes before decreasing again after $T_E > 35$mins. This flat region is interesting as it represents a region upon which the network structure is invariant with T_E, hence the spreading dynamics is maintained for this region. The network slowly becomes disconnected due to the removal of many of the edges with T_E below the threshold. The size of the GC reduces rapidly to half of the original number after a threshold of 15mins is applied. (refer to Fig. 3) The total number of components thus increases accordingly due to the increasing number of nodes which become disconnected from the GC. Although the size of the GC decreases, its diameter increases gradually from 7 to a maximum of 41 at 50mins. The mean component size drops rapidly at $T_E = 1$ to approximately 1, indicating that most of the nodes essentially become isolated from each other instead of forming smaller clusters (i.e. many of these nodes have 0 incoming edges).

In summary, we observe that the network rapidly loses its ability to infect the full global population as we increase the threshold to about 15-20mins since a majority of the commuters are no longer connected to each other. However, we note that even after a high threshold is applied, the GC endures and does not abruptly fracture into smaller components. This is unlike the abrupt phase change behaviour observed that resembles percolation in random networks [1].

5 Structural attack and random failure of network structure

We analyse the resilience of the bus encounter network by gradually removing a percentage of nodes and the edges associated with them. Three different strategies are compared to each other; a) Random removal of a selected percentage of nodes, b) targeted removal of highest degree nodes first, and c) Removal of highest betweenness centrality first where betweenness centrality is defined as:

$$b_c(v) = \sum_{s,t \in V, s \neq t} \frac{\sigma(s \to t|v)}{\sigma(s \to t)} \tag{2}$$

where $\sigma(s \to t)$ is the sum of shortest paths between the nodes s and t while $\sigma(s \to t|v)$ is the sum of shortest paths between s and t that pass through the node v. We use this

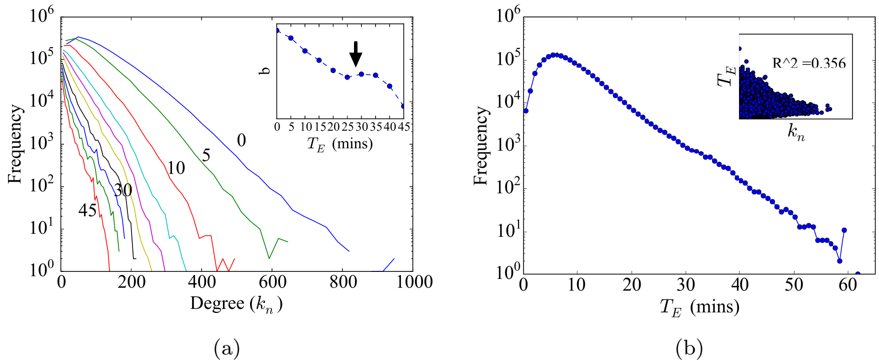


Figure 2: Distribution curves for degree and mean encounter duration respectively. (a) The degree distribution of the encounter network closely resembles an exponential distribution $f(k_n) \approx a \exp^{bk_n}$ even after successive thresholding is applied from 0 to 50mins. The numbered labels represent the amount of threshold (in mins) applied to the network. The inset plot shows the variation in exponent b of the degree distribution curves after linear fitting. The arrow highlights the range of T_E for which b remains relatively invariant. (b) The encounter duration distribution also resembles an exponential distribution although when plotted against degree in the inset, it reveals little correlation between the two quantities.

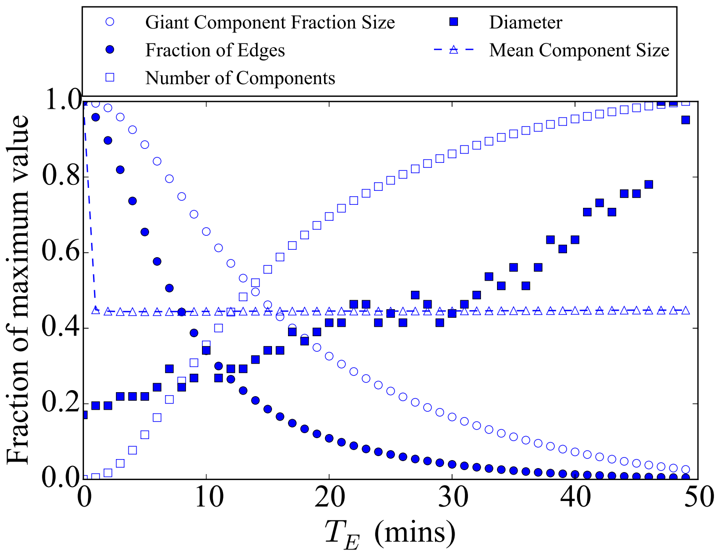


Figure 3: The plot of different network properties as the threshold encounter duration is increased. The values are plotted here as a fraction of their maximum obtained values so that they may be plotted on the same axis. For reference and completeness, we provide the maximum values here: the maximum mean component size (△) obtained at T_E = 0mins is 2.26 which drops rapidly to 1 at T_E =1 min. The maximum diameter (■) and the maximum number of components (□) obtained at T_E=50mins is 41 and 1,434,580 respectively.

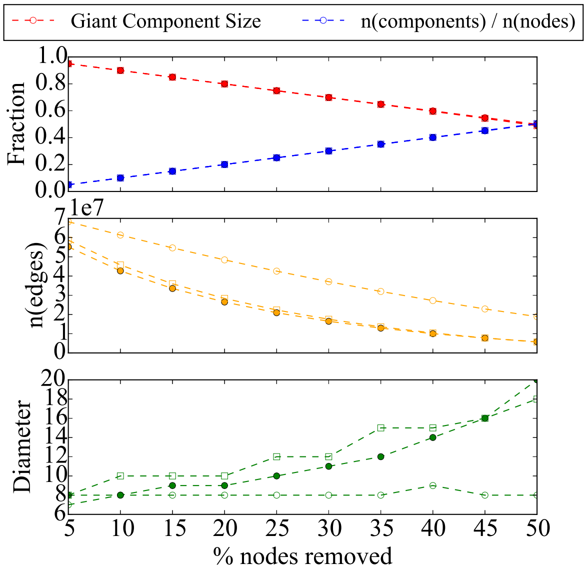


Figure 4: The plot of different network properties varying with the number of nodes removed in discrete steps of 5% from 5% to 50%. Unfilled circles (○) represent randomly removed nodes, filled circles(●) targeted removal of highest degree nodes while square markers (□) represent removal of nodes with highest betweenness. Dashed lines represent connecting guides for the eye.

third strategy as an alternative method of attack because evidence from the subset network visualizations suggested that several low degree nodes formed bridges between different journey networks (see Fig. 1(b)). We conjectured that these nodes could in fact be identified by their higher betweenness centrality scores.

Since the network shows similar degree structure to a random network, we conjecture that targeted removal of highest degree nodes might not be as effective in breaking the network structure as compared to the removal of nodes in a scale-free network. The results support the idea as the method of removal did not affect the rate of change in size of GC or the number of components (see Fig. 4) despite the fact that an overall greater number of edges are removed for the cases of targeted removal as compared to random removal. However, the results did indicate that targeted removal significantly increases the diameter of the GC as compared to random removal. The difference between the removal of high betweenness nodes as compared to high degree is not significant. Since the calculation of betweenness centrality is computationally more expensive as compared to node degree, it is therefore not a good choice for targeted removal of nodes in this network.

6 Simulating Epidemic Spread in the Encounter Network

Finally, we test the capability of the network when subjected to several infected commuters using the standard SIR epidemic spread model. As our network is based on the encounters in a single day, we represent each day as one time step in our simulation. All the commuters begin in the susceptible (S) state and a small number of index commuters are placed into the infected state (I). Commuters adjacent to these index cases in the network have a probability β of being infected by them and will transition into the infected state (I) themselves during the next time step. After being in the infected state, the commuter will have a probability γ of transitioning into the final recovered (R) state where the commuter is now immune to the disease in question (see Fig 5(a)).

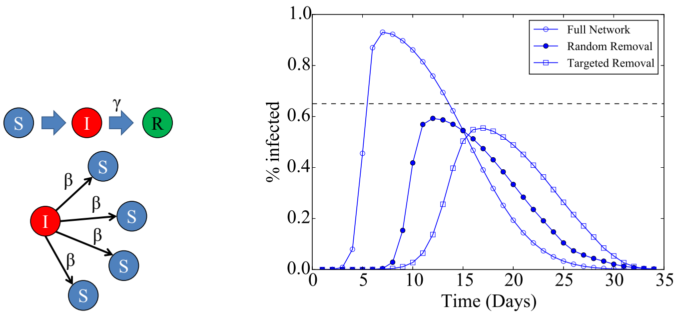
We make several assumptions in our model; the network is assumed to remains static throughout the period of the simulation, disregarding the effect of the weekends which would have created different network structures due to the changes in the amount and pattern of travel for the commuters. We also assume that infection happens solely through the bus encounter network, ignoring the effect of other mechanisms of contact such as those occurring either at home or at the workplace. We also assume that after getting infected, the commuter travel patterns do not change which, in reality does change due to commuters being more likely to stay at home or going to the see the doctor whilst refraining from going to work or school. We also set the simulation such that newly infected passengers may only infect other passengers on the day after being infected. This mechanism therefore provides an automatic incubation period of approximately one day between the time that commuter gets infected to the point where he himself is infectious to other commuters.

Our selection for the appropriate infection and recovery rates β and γ is based on the infection and recovery rates reported in the simulation conducted for a high resolution contact network located around a school[13]. This in turn was based on the infection rate of influenza spreading through a commercial airliner [9] which is equivalent to an infection rate of 0.003 for every 20 seconds. For the purpose of our simulation which has a time step of 1 day, we simplify by using the mean encounter duration of 9.64mins to calculate a flat infection rate per day for all commuter interactions. This provides us with the final selected infection rate of $\beta_{1day} = 1-(1-\beta_{20sec})^{\frac{9.64\times 60sec}{20sec}} = 0.083$. The time taken for commuters in our model to recover from the disease is normally distributed with mean $1/\gamma = 10$ days with a variance of 5 days while the initial number of infected individuals consisted of 10 randomly chosen commuters.

The results show that within 7 days of the simulation, the number of infected commuters rises rapidly and exponentially up to a maximum of 93.03% out of all the commuters. (see Fig. 5(b)) The network commuters recover from the epidemic ($n(I) = 0$) only after about 30 days. We compare this result to against two different simulations in which the network is subjected to edge removal using both targeted and random removal strategies. It is shown that amongst the nodes that remained in the network, a maximum of about 90% of the remaining nodes are still infected with little difference between the two strategies. However, for the case of targeted removal, we observed that the time of maximum infection is delayed by about 5 days. This evidence therefore supports our earlier statement where we showed that while the connectedness of the network is similar for different removal strategies, the differences in diameter of the resulting GC affected the speed in which the disease is transmitted through the network.

7 Conclusion and discussion

In summary, we have constructed the network generated from encounters between commuters sharing the same public bus for a typical weekday for the entire population of commuters in a typical metropolitan city. The network has been shown to have an immense capability of transmitting diseases due to its unique structure which is highly clustered, extremely dense and resistant to targeted node removal. The network is also well connected with 99.97% of the commuters within 7 steps of each other, and our simulations reveal that the network has the



(a) (b)

Figure 5: a) A schematic diagram of the standard SIR model with three states: susceptible (S), infected (I) and recovered (R). b) The dynamics of growth of infected commuters in the encounter network with an infection rate $\beta = 0.083$ and the recovery rate of $\gamma = 0.1$ per day. Three scenarios are tested: The full network ($\circ$), a network where 35% of randomly selected nodes are isolated ($\bullet$) and network where 35% of highest degree nodes are isolated ($\square$). The horizontal blue line shows the maximum % of nodes (65%) still connected to the network relevant to the two removal cases.

ability to infect the entire population of commuters within 7 days of the onset of infection. While these initial results may be alarming, we show that when a threshold on the encounter duration is applied, the network fragments easily. This means that only highly infectious diseases which do not require much time (<10mins) for transmission are actually capable of infecting the entire network. Although targeted removal of essential nodes is found to be unable to isolate the bulk of nodes found within the remaining giant component, it is also shown to be able to delay the onset of the epidemic thus other different combinations of mitigation strategies could be more easily employed.

Our study indicates that the mechanism of spread of diseases during the encounters of commuters with other strangers whilst using the public transportation system is critical and should not be ignored in a comprehensive model. While contact network tracing could be easily conducted for individuals sharing the same workplace or home, it is much harder for one to trace the spread of disease for highly dynamic networks such as the one highlighted in this article. The availability of travel data, such as that used in this study should therefore be an essential portion of future contact network tracing for a highly infectious outbreak. This work also largely ignores the significant role of bus drivers who will certainly pose as massive hubs for infection spreading within this commuter network. If data which bus drivers were on duty could be made available, the impact of bus drivers as compared to commuter could be modelled and quantified, and different strategies for targeted removal could therefore be devised. Our work also paves the way for more detailed analyses hopefully resulting from the combination of travel datasets such as that identified in this work together with datasets identifying the home and workplaces/schools of the commuters.

7.1 Acknowledgments

We would like to thank the Land Transport Authority of Singapore for providing us the smart-card commuter ridership data that is used in this study to build the massive commuter encounter network. This research is supported by the Science and Engineering Research Council (SERC) of the Agency for Science, Technology and Research (A*STAR) of Singapore. (Complex Systems Programme grant number 122 45 04056).

References

[1] Reka Albert, Hawoong Jeong, and Albert-László Barabási. Attack and error tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[2] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.

[3] Duygu Balcan and Alessandro Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Physics*, 7(7):581–586, 2011.

[4] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. In *Proc. International AAAI Conference on Weblogs and Social Media*, 2009.

[5] Duan-Bing Chen, Rui Xiao, and An Zeng. Predicting the evolution of spreading on complex networks. *Scientific Reports*, 4:06108, 2014.

[6] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[7] M Marder. Dynamics of epidemics on random networks. *Physical Review E*, 75(6):066103, 2007.

[8] Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports*, 1, 2011.

[9] Michael R Moser, Thomas R Bender, Harold S Margolis, Gary R Noble, Alan P Kendal, and Donald G Ritter. An outbreak of influenza aboard a commercial airliner. *American Journal of Epidemiology*, 110(1):1–6, 1979.

[10] Land Transport Authority of Singapore. Mytransport.sg - commuting - bus services. Accessed [online] on:, January 2015. `http://www.mytransport.sg/content/mytransport/home/commuting/busservices.html`.

[11] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200, 2001.

[12] Sen Pei, Lev Muchnik, José S Andrade Jr, Zhiming Zheng, and Hernán A Makse. Searching for superspreaders of information in real-world social media. *Scientific Reports*, 4:05547, 2014.

[13] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.

[14] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Manuel Cebrian. Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific Reports*, 4, 2014.

[15] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.

[16] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[17] Kenta Yashima and Akira Sasaki. Epidemic process over the commute network in a metropolitan area. *PLOS ONE*, 9(6):e98518, 2014.